



①9 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

⑫ Patentschrift  
⑩ DE 101 22 212 C 1

⑤1 Int. Cl.<sup>7</sup>:  
G 10 L 15/06  
G 10 L 15/14  
G 06 K 9/66

②1 Aktenzeichen: 101 22 212.2-53  
②2 Anmeldetag: 8. 5. 2001  
④3 Offenlegungstag: –  
④5 Veröffentlichungstag  
der Patenterteilung: 10. 10. 2002

DE 101 22 212 C 1

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

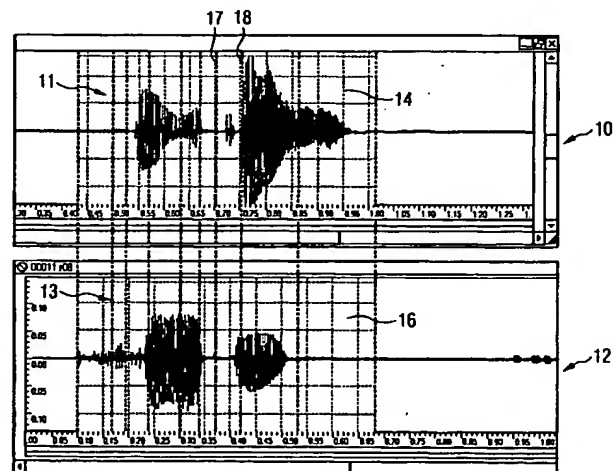
⑦3 Patentinhaber:  
Siemens AG, 80333 München, DE

⑦2 Erfinder:  
Sterz, Walter, Wien, AT; Tschirk, Wolfgang, Wien, AT

⑤6 Für die Beurteilung der Patentfähigkeit in Betracht  
gezogene Druckschriften:  
DE 197 08 183 A1  
US 54 73 728

⑤4 Verfahren zum Trainieren eines Modells für die Mustererkennung, insbesondere eines  
Hidden-Markov-Modells, und Mustererkenner

⑤7 Die Erfindung betrifft ein Verfahren zum Trainieren ei-  
nes Modells für die Spracherkennung, insbesondere ei-  
nes Hidden-Markov-Modells, bei dem a) ein Verschie-  
bungswert zwischen mindestens zwei zusammengehö-  
renden Trainingsmustern (11, 13) durch Korrelieren derart  
bestimmt wird, daß die Korrelation zwischen den Train-  
ingsmustern maximal wird, und b) die Trainingsmuster  
(11, 13) mittels des Verschiebungswertes derart zueinan-  
der verschoben werden, daß einander entsprechende Be-  
reiche der Trainingsmuster bezüglich eines gemeinsamen  
Bezugspunktes etwa die gleiche Lage aufweisen.



DE 101 22 212 C 1

[0001] Die Erfindung betrifft ein Verfahren zum Trainieren eines Modells für die Mustererkennung gemäß Anspruch 1 und einen Mustererkenner gemäß Anspruch 10.

5 [0002] Mustererkenner werden auf vielen Gebieten eingesetzt: Spracherkennung, Bildverarbeitung, ökologische und ökonomische Prozesse. Insbesondere für die Spracherkennung auf den Gebieten der Telekommunikation, in Umgebungssteuerungen (beispielsweise der Home-Automation) und Gerätesteuerungen kommen Mustererkenner auf der Basis des Hidden-Markov-Modells oder von neuronalen Netzen zum Einsatz.

10 [0003] Bei der automatischen Mustererkennung werden aus einem ein Muster repräsentierenden Signal Merkmalsvektoren extrahiert. Beispielsweise werden hierzu Abschnitte eines Sprachsignals jeweils mittels Fouriertransformation vom Zeit- in den Frequenzbereich transformiert, so daß die Merkmalsvektoren die spektralen Anteile der einzelnen Abschnitte repräsentieren. Für ein ein Muster repräsentierendes Signal erhält man so eine Abfolge von Merkmalsvektoren, die spektrale Anteile, Energieanteile, etc. des Signals repräsentieren. Die Vektoren gehören zu einem hochdimensionalen Merkmalsraum.

15 [0004] Bei einem Sprachsignal sind in einem Merkmalsraum einzelnen Phonemen bzw. Lauten bestimmte Raum-Bereiche zugeordnet. Anhand der zeitlichen Abfolge von Merkmalsvektoren eines Sprachsignals in dem hochdimensionalen Merkmalsraum und der Lage einzelner Merkmalsvektoren in Bezug auf die den einzelnen Phonemen zugeordneten Raum-Bereiche ist es möglich, einen zu dem Sprachsignal passenden Text zu ermitteln oder aufgrund des erkannten Sprachsignals Steuervorgänge auszuführen.

20 [0005] Aus der DE 197 08 183 A1 ist ein Verfahren zur Spracherkennung mit Sprachmodellanpassung an die Besonderheiten eines speziellen Textes bekannt. Hierbei werden aus einem Sprachsignal Testwerte abgeleitet und mit ein vorgegebenes Vokabular bestimmenden Referenzwerten verglichen. Hieraus werden Bewertungswerte abgeleitet, die an Wortgrenzen mit Sprachmodellwerten verknüpft werden, welche von der Wahrscheinlichkeit abhängen, dass ein bestimmtes Wort des Vokabulars in Abhängigkeit von wenigstens einem vorhergehenden Wort auftritt. Im Rahmen dieses

25 Verfahrens werden sogenannte Unigramm- und M-Gramm-Basis Sprachmodellwerte sequentiell eingesetzt. [0006] Die Zuordnung einer Folge von Merkmalsvektoren zu einem bestimmten Muster kann mittels eines Hidden-Markov-Modells (HMM) oder eines neuronalen Netzes erfolgen.

[0007] Markov-Modelle sind aus der statistischen Signaltheorie bekannt und beschreiben mit einer besonders übersichtlichen Struktur statistische Prozesse dadurch, daß anhand eines Zustandsübergangsgraphen die Übergangswahrscheinlichkeiten zwischen verschiedenen Zuständen eines Systems dargestellt werden. Im Bereich der automatischen Mustererkennung hat sich das Hidden-Markov-Modell zur Zuordnung von Merkmalsvektorfolgen im hochdimensionalen Merkmalsraum zu einem bestimmten Muster bewährt, das mit höchster Wahrscheinlichkeit dem in einem Signal enthaltenen Muster entspricht.

30 [0008] Bei einem Spracherkenner bildet das Hidden-Markov-Modell für jedes Phonem oder Einzelwort einen Zustandsgraphen mit einer den Phonemsegmenten entsprechenden Anzahl von Zuständen. Übergangswahrscheinlichkeiten zwischen den verschiedenen Zuständen ermöglichen die Modellierung von Phonemsegmentfolgen mit unterschiedlichen Wahrscheinlichkeiten. Ausgehend von einem vorliegenden, ein Muster repräsentierenden Signal erhält man durch Signalvorverarbeitung jedoch nur den Weg durch den hochdimensionalen Merkmalsraum, nicht aber die Folge von Raum-Bereichen (entsprechend einer Folge von Phonemen bei einem Sprachsignal).

40 [0009] Vor dem Einsatz eines Hidden-Markov-Modells zur Mustererkennung ist ein Trainieren erforderlich, um überhaupt die Raum-Bereiche in dem hochdimensionalen Merkmalsraum zu erzeugen. Trainiert wird beispielsweise bei einem Spracherkenner durch das Sprechen von Trainingsmustern, die bereits bekannt sind. Durch das Trainieren werden Übergangswahrscheinlichkeiten im Hidden-Markov-Modell eingestellt sowie die Verteilung der Trainingsmuster in einem Segment modelliert, um eine möglichst exakte Erkennung zu erzielen. Als Trainingsmuster werden entweder einzelne Worte oder gesprochene Sätze verwendet.

45 [0010] Aus der US 5,473,728 ist ein Verfahren zum Trainieren eines sogenannten homoscedastischen Hidden-Markov-Modells zur automatischen Spracherkennung bekannt, welches die Schritte der Bereitstellung einer akustischen Spracheinheiten enthaltenden Datenbasis, der Erzeugung des besagten Hidden-Markov-Modells aus den akustischen Spracheinheiten und das Laden dieses Modells in den Spracherkenner umfasst. In einer vorteilhaften Ausprägung dieses Verfahrens werden mehrere iterative Trainingsschritte ausgeführt, wobei jeweils eine Vorwärts-Zustandswahrscheinlichkeit für einen Markov-Kettenzustand, eine Rückwärts-Zustandswahrscheinlichkeit für den Markov-Kettenzustand und eine Komponenten-Zustandswahrscheinlichkeit für den Markov-Kettenzustand sowie eine Misch-Gauss-PDF-Komponente rekursiv für einen gegebenen Messungsvektor, eine Trainingssequenz und eine Spracheinheit berechnet werden. Im Abschnitt "Background of the Invention" dieser Druckschrift wird auf eine Reihe von Veröffentlichungen zur Handhabung von Hidden-Markov-Modellen, speziell unter dem Gesichtspunkt ihres Einsatzes bei der Spracherkennung, hingewiesen.

55 [0011] Besonders wichtig für die Erstellung und das Trainieren eines Modells für die Mustererkennung, insbesondere eines Hidden-Markov-Modells, ist die Segmentierung der verwendeten Trainingsmuster. Darunter versteht man die Unterteilung in bestimmte Abschnitte, in denen die Signalparameter annähernd konstant bleiben, beispielsweise bei einem Sprachsignal als Trainingsmuster die Unterteilung in Abschnitte mit gleichem Lautinhalt.

60 [0012] Allerdings sind nicht immer in allen Trainingsmustern alle Ausprägungen eines Signals vorhanden. Beispielsweise können Laute am Anfang oder Ende eines Wortes fehlen. Dies ist insbesondere dann kritisch, wenn zusammengehörende Trainingsmuster, beispielsweise ein bestimmtes Wort, während des Trainings mehrfach auftauchen. Geht man von den Muster Grenzen eines Trainingsmusters aus und wird innerhalb dieser Grenzen segmentiert, kann es vorkommen, daß Segmente, denen der gleiche Index aufgrund eines zusammengehörenden Trainingsmusters zugeordnet ist, Bereiche unterschiedlichen Lautinhalts repräsentieren.

65 [0013] In Fig. 3 ist im oberen Diagramm 10 der zeitliche Verlauf eines ersten, das Wort "Sieben" (ausgesprochen als "Sieben") repräsentierenden Sprachsignals 11 dargestellt. Im unteren Diagramm 12 ist der zeitliche Verlauf eines zweiten, ebenfalls das Wort "Sieben" (ausgesprochen allerdings als "Siebn") repräsentierenden Sprachsignals 13 abgebildet.

Die hell dargestellten Bereiche 14 und 16 entsprechen den akustisch ermittelten und vom Sprachsignal überdeckten Bereichen, welche die Wortgrenzen vorgeben. Die Bereiche unterscheiden sich erkennbar für beide Sprachsignale. Dadurch führt hier eine an den Wort- bzw. Mustangrenzen ausgerichtete Segmentierung und anschließende Indizierung dazu, daß in beiden Trainingsmustern Segmente mit gleichem Index einen unterschiedlichen Lautinhalt aufweisen.

[0014] Bei der Segmentierung wird bisher von einer Initialsegmentierung ausgegangen, die im einfachsten Fall als äquidistant angesetzt wird. Konkret bedeutet dies, daß ein Trainingsmuster zur Segmentierung in äquidistante Abschnitte eingeteilt wird. Die einzelnen Grenzen von Segmenten können jedoch mittels Viterbi-Segmentierung entsprechend dem Verlauf eines Trainingsmusters im Hidden-Markov-Modell bei einer Erkennungsphase verschoben werden.

[0015] Ferner sind zum Trainieren sogenannte "geschnittene", d. h. zeitlich begrenzte Trainingsmuster erforderlich. In einem dem Training vorhergehenden Schritt werden daher für jedes einzelne Trainingsmuster sogenannte Mustangrenzen bestimmt (siehe die Bereiche 14 und 15 in Fig. 3). Dies bedeutet, daß das Trainingsmuster zeitlich begrenzt oder anders ausgedrückt isoliert sein muß, wenn die Ermittlung der Mustangrenzen automatisiert werden soll. Liegt jedoch als Trainingsmuster ein kontinuierliches Signal vor, ist eine Automatisierung der Ermittlung der Mustangrenzen so gut wie nicht möglich.

[0016] Der vorliegenden Erfindung liegt daher die Aufgabe zugrunde, ein Verfahren zum Trainieren eines Modells für die Mustererkennung, insbesondere eines Hidden-Markov-Modells, und einen Mustererkenner vorzuschlagen, die eine im Vergleich zu einer "starrten" Initialsegmentierung verbesserte Segmentierung ermöglichen.

[0017] Diese Aufgabe wird durch ein Verfahren mit den Merkmalen von Anspruch 1 und durch einen Mustererkenner mit den Merkmalen von Anspruch 10 gelöst. Bevorzugte Ausgestaltungen des Verfahrens und des Mustererkenners ergeben sich aus den jeweiligen abhängigen Ansprüchen.

[0018] Die Erfindung schließt den grundlegenden Gedanken ein, ein Trainieren eines Modells für die Mustererkennung mittels mehrerer gleichartiger Trainingsmuster, insbesondere eines Hidden-Markov-Modells, vorerst ohne explizite Bestimmung von Mustangrenzen zu ermöglichen, indem ein Verschiebungswert zwischen mindestens zwei zusammengehörenden Trainingsmustern durch Korrelieren der Trainingsmuster bestimmt wird. Der Verschiebungswert gibt an, wie ein Trainingsmuster in Bezug auf ein anderes Trainingsmuster verschoben werden muß, damit Bereiche mit gleichem Inhalt in den Trainingsmustern an gleicher Stelle liegen. Hierdurch ist keine explizite Bestimmung von Mustangrenzen erforderlich.

[0019] Verfahrensgemäß wird a) ein Verschiebungswert zwischen mindestens zwei zusammengehörenden Trainingsmustern durch Korrelieren derart bestimmt, daß die Korrelation zwischen den Trainingsmustern maximal wird. Dann werden b) die Trainingsmuster mittels des Verschiebungswertes derart zueinander verschoben, daß einander entsprechende Bereiche der Trainingsmuster bezüglich eines gemeinsamen Bezugspunktes etwa die gleiche Lage aufweisen. Einzelne Grenzen für die Trainingsmuster sind hierzu nicht notwendig.

[0020] Vorzugsweise kann das Korrelieren iterativ über mehrere Durchläufe für die Trainingsmuster durchgeführt werden, wodurch der Verschiebungswert besonders exakt ermittelt wird. Dies ist für eine besonders hohe Erkennungsrate wichtig, da hierdurch mit hoher Wahrscheinlichkeit Segmenten im Trainingsmuster, die den gleichen Inhalt aufweisen, auch derselbe Index zugeordnet wird. In einer konkreten Ausführungsform werden die zueinander verschobenen Matrizen von zwei Trainingsmustern addiert und bilden somit eine Mittelwertmatrix. Die Verschiebung aller weiteren Trainingsmuster wird durch Korrelation mit dieser Mittelwertmatrix bestimmt, wobei die neuerlich korrelierten Trainingsmuster wiederum verschoben zur Mittelwertmatrix dazu addiert werden. Dieser Vorgang kann iterativ mehrmals über alle Trainingsmuster erfolgen.

[0021] Vorzugsweise werden in einem Trainingsmuster Bereiche, die im Vergleich mit anderen zusammengehörenden Trainingsmustern fehlen, durch vorgegebene Füllbereiche ergänzt. Diese vorgegebenen Füllbereiche können beispielsweise bei einem Sprachsignal lauffreie Bereiche (silence) sein. Insbesondere wird dies bei Trainingsmustern vorgenommen, bei denen beispielsweise Laute am Anfang oder am Ende eines Wortes fehlen.

[0022] In einer bevorzugten Ausgestaltung des Verfahrens werden anhand der in Schritt b) vorgenommenen Verschiebung Mustangrenzen bestimmt, die für alle zusammengehörenden Trainingsmuster gelten. Erfolgt danach eine Segmentierung, die sich auf diese Mustangrenzen bezieht, werden verschiedene, aber zusammengehörende Trainingsmuster derart segmentiert, daß Bereiche mit gleichem Inhalt in den verschiedenen Trainingsmustern in den gleichen Segmenten liegen.

[0023] Ein zwischen den Mustangrenzen liegender Bereich der Trainingsmuster kann mittels eines Zentroidenverfahrens segmentiert werden. Gemäß dem Zentroidenverfahren können innerhalb vorgegebener Beschränkungen alle möglichen Segmentierungen berechnet werden. Aus den berechneten Segmentierungen wird dann eine Segmentierung nach einem Distanzkriterium zum Zentroiden jedes Segments ausgewählt.

[0024] Das Auswahlkriterium kann derart ausgebildet sein, daß die Segmentierung ausgewählt wird, bei der die Summe der Quadrate der Distanzen von Merkmalsvektoren zu dem zugeordneten Zentroiden in jedem Segment über das gesamte Trainingsmuster minimal wird. Hierdurch erhält man eine optimale Segmentierung.

[0025] Schließlich kann die Segmentierung entweder direkt zum Erstellen eines Modells zur Mustererkennung, insbesondere eines Hidden-Markov-Modells, oder als Initialsegmentierung erfolgen. Anhand der Initialsegmentierung kann die Verschiebung von Segmentgrenzen einzelner Trainingsmuster, insbesondere mittels Viterbi-Segmentierung, vorgenommen werden.

[0026] Vorzugsweise wird das Verfahren bei der Spracherkennung eingesetzt.

[0027] Ein Mustererkenner zur Durchführung des erfindungsgemäßen Verfahrens weist

- Vorverarbeitungsmittel zum Vorverarbeiten eines ein Trainingsmuster repräsentierendes Signal,
- Berechnungsmittel zum Erstellen von Matrizen mit Merkmalsvektoren für das Trainingsmuster,
- Korrelationsmittel zum Korrelieren von Matrizen, die zusammengehörende Trainingsmuster repräsentieren,
- Auswertemittel zum Ermitteln eines Verschiebungswertes, und
- Sortiermittel zum Umsortieren von Matrizen abhängig von dem ermittelten Verschiebungswert auf.

[0028] Ferner können Additionsmittel vorgesehen sein, welche die Matrizen nach einem Umsortieren addieren.

[0029] Weitere Vorteile und Anwendungsmöglichkeiten der Erfindung ergeben sich aus der nachfolgenden Beschreibung eines Ausführungsbeispiels des erfindungsgemäßen Verfahrens in Verbindung mit den Zeichnungen. In diesen ze-

5 gen  
[0030] Fig. 1 die in Fig. 3 dargestellten Trainingsmuster nach einer Verschiebung entsprechend einem mit dem erfindungsgemäßen Verfahren ermittelten Verschiebungswert,

[0031] Fig. 2 eine Segmentierung mit neun Segmenten der in Fig. 1 dargestellten Trainingsmuster, und

[0032] Fig. 3 ein Beispiel von zwei zusammengehörenden Trainingsmustern, die zueinander verschoben sind.

10 [0033] Zur Beschreibung von Fig. 3 wird auf die Beschreibungseinleitung verwiesen.

[0034] Ein Trainingsmuster für einen Spracherkenner wird durch eine Matrix dargestellt, deren Spalten Merkmalsvektoren mit aufeinanderfolgenden Zeitindizes entsprechen. Hierbei ist die Zeit in sogenannte Frames als kleinste Zeiteinheit unterteilt. Ein Frame kann beispielsweise eine bestimmte Anzahl von digitalisierten Abtastwerten eines analogen Signals sein, welches das Trainingsmuster repräsentiert. Jeder Zeitindex entspricht einem Vielfachen einer Framelänge.

15 [0035] Die Zeilen der Matrix stellen Merkmale dar, welche durch die Vorverarbeitung des Signals gewählt wurden. Als Beispiele seien hier nur spektrale Inhalte, Energiewerte, etc. genannt.

[0036] Im folgenden wird mit i ein Frame-Index, mit j ein Segment-Index und mit k eine Merkmals-Komponente bezeichnet.

20 [0037] Durch Korrelieren zweier zusammengehörender Trainingsmuster, die beispielsweise ein bestimmtes Wort in einem Sprachsignal repräsentieren, wird ein Zeitverschiebungswert bestimmt. Der Zeitverschiebungswert gibt die zeitliche Verschiebung der Trainingsmuster, bezogen auf einen gemeinsamen Bezugspunkt der beiden Trainingsmuster, an. Er wird durch das Korrelieren derart berechnet, daß die Korrelation maximal wird. Sind mit f und g zwei Matrizen bezeichnet, die jeweils ein Trainingsmuster repräsentieren, die beide zusammengehören, und bedeutet m die Anzahl der Merkmalsvektoren pro Äußerung, so ergibt sich der Zeitverschiebungswert nach der folgenden Gleichung:

$$25 \quad c_{f,g}(x) = \sum_{i=0}^{m-1} \sum_k^k f_i g_{(i+x) \bmod m}^k$$

30 mit

$$x = \arg \max \{ c_{f,g}(x) \}$$

35 [0038] Konkret gibt der Zeitverschiebungswert x hier an, um wieviele Frames ein Trainingsmuster bezüglich des anderen Trainingsmusters verschoben werden muß, so daß entsprechende Bereiche der beiden Muster "zeitlich zur Deckung kommen". Beginnt beispielsweise eine Äußerung in einem ersten Trainingsmuster mit dem Merkmalsvektor 0 (= erste Spalte der entsprechenden Matrix) und ergibt sich als Verschiebungswert eines zweiten Trainingsmusters bezüglich des ersten Trainingsmusters 10, so heißt das, daß die Äußerung im zweiten Trainingsmuster mit dem Merkmalsvektor 10  
40 entsprechend der elften Spalte der Matrix beginnt. Um Merkmalsvektoren, die zu einander entsprechenden Bereichen der Trainingsmuster gehören, zur "Deckung" zu bringen, können die Spalten der zweiten Matrix jeweils um zehn "Plätze" verschoben werden.

[0039] Die Korrelation kann iterativ über mehrere Durchläufe durchgeführt werden, um einen exakten Zeitverschiebungswert zu erhalten.

45 [0040] Vorerst muß der Bereich vor und hinter jedem Trainingsmuster nicht weggeschnitten werden. Mustergrenzen sind daher nicht notwendig.

[0041] In Fig. 1 sind die in Fig. 3 zeitlich verschobenen Trainingsmuster 11 und 13 in dem oberen und unterem Diagramm 10 bzw. 12 nach einer Zeitverschiebung des zweiten Trainingsmusters 13 gemäß dem vorgenannten Verfahren dargestellt. Die einander entsprechenden Bereiche der beiden Trainingsmuster 11 und 13 liegen nun zeitlich an gleichen  
50 Positionen. Die durch die hellen Bereiche dargestellten Bereiche 14 und 16 der beiden Trainingsmuster 11 und 13 sind mit ihren Wortgrenzen nun aneinander angepaßt.

[0042] Um einen Mittelwert mehrerer Trainingsmuster zu erhalten, werden nach der zeitlichen Verschiebung die Trainingsmuster addiert, d. h. die beiden die Trainingsmuster repräsentierenden Matrizen f und g werden zu einer neuen Matrix h addiert. Ein drittes Trainingsmuster i, das ebenfalls zu den beiden Trainingsmustern gehört, kann anschließend mit dem oben erläuterten Verfahren zeitlich in Bezug auf die Matrix h entsprechend dem kumulierten Trainingsmuster "korrigiert und wiederum zur Matrix h addiert werden. Das "Korrigieren" umfaßt das Korrelieren der Matrizen h und i und Ermitteln des Zeitverschiebungswertes x sowie das Umordnen der Spalten der Matrix i mit dem Zeitverschiebungswert x. Insgesamt erhält man so einen Mittelwert (oder genauer gesagt eine Mittelwert-Matrix) von mehreren zusammengehörenden Trainingsmustern.

60 [0043] Mit dem Mittelwert über mehrere Trainingsmuster können nun gemeinsame Wortgrenzen bestimmt werden, die für alle zusammengehörenden Trainingsmuster gelten. Dies ist bereits in Fig. 1 dargestellt: Die durch die hellen Bereiche 14 und 16 dargestellten Wortgrenzen sind aneinander angepaßt. Oder anders ausgedrückt: Die Wortgrenzen werden durch das Korrelieren von Trainingsmustern und anschließendes Verschieben oder Aneinander-Ausrichten der Trainingsmuster insgesamt aneinander angepaßt. Die so ermittelten und allen Trainingsmustern gemeinsamen Wortgrenzen entsprechen den äußersten Wortgrenzen aller korrelierten und verschobenen Trainingsmuster. Hierdurch führt eine  
65 anschließende Segmentierung der Trainingsmuster dazu, daß in gleichen Segmenten verschiedener Trainingsmuster jeweils Bereiche mit gleichem Lautinhalt liegen.

[0044] In einzelnen Trainingsmustern fehlende Laute werden hierbei durch Bereiche ersetzt, die lautmäßig "Stille" ent-

sprechen (sogenannte "Silence"-Bereiche).

[0045] Innerhalb der gemeinsamen Wortgrenzen kann nun eine Segmentierung durchgeführt werden. Die Segmentierung basiert auf einem Zentroidenverfahren. Gemäß dem Zentroidenverfahren werden innerhalb vorgegebener Beschränkungen alle möglichen Segmentierungen berechnet und daraus die beste nach dem Distanzkriterium zum Zentroiden ausgewählt.

[0046] Es seien  $f_i$  ein Merkmalsvektor der Mittelwertmatrix  $h$  mit Frame-Index  $i$  und  $u_j, v_j$  Frame-Indices der Segmentgrenzen. Der Zentroid  $z_j$  ist dann der Mittelwertsvektor eines Segments  $j$ :

$$z_j = \frac{1}{v_j - u_j} \sum_{i=u_j}^{v_j-1} f_i \quad 10$$

[0047] Die Segmentierung ist optimal, wenn die Distanzsumme der Merkmalsvektoren zu ihrem Zentroiden im Segment über das gesamte Trainingsmuster minimal wird.

[0048] Als Beschränkungen seien folgende Werte angenommen:

$n_{\min}$  minimale Anzahl Merkmalsvektoren pro Segment,

$n_{\max}$  maximale Anzahl Merkmalsvektoren pro Segment,

$N$  Anzahl Segmente,

$L$  Anzahl Merkmalsvektoren pro Muster (ist für alle Muster gleich).

[0049] Mit der folgenden Gleichung werden die Distanzsummen der Merkmalsvektoren zu ihrem Zentroiden über alle möglichen Segmentierungen berechnet:

$$D(u_j, v_j) = \sum_{j=1}^N \sum_{i=u_j}^{v_j-1} \|z_j - f_i\|^2 \quad 25$$

[0050] Die Segmentgrenzen  $(u_j, v_j)$  werden innerhalb der obigen Beschränkungen nach den folgenden Kriterien variiert:

$$(u_j, v_j) = \arg \min \{D(u_j, v_j)\} \quad \text{mit} \quad n_{\min} \leq v_j - u_j \leq n_{\max} \quad 30$$

$$\text{und} \quad \sum_{j=1}^N (v_j - u_j) = L$$

[0051] Hierbei werden alle möglichen Kombinationen durchgerechnet, was zum globalen Optimum führt. Die so ermittelte Segmentierung kann direkt zum Erstellen eines Hidden-Markov-Modells herangezogen werden oder als Initialsegmentierung für die Verschiebung der Segmentgrenzen der einzelnen Trainingsmuster beispielsweise mittels Viterbi-Segmentierung während des Trainings dienen.

[0052] In Fig. 2 ist eine Segmentierung der Trainingsmuster aus Fig. 1 nach dem Zentroidenverfahren dargestellt. Die beiden Trainingsmuster wurden, ausgehend von den Mustergerenzen, in neun Segmente eingeteilt. Jedes Segment weist Bereiche der beiden Trainingsmuster auf, die den gleichen Lautinhalt umfassen. Die Segmentgrenzen sind mittels gepunkteter Linien in dem oberen und unteren Diagramm 10 und 12 dargestellt. Das siebte Segment weist beispielsweise die Segmentgrenzen 17 und 18 auf. Erkennbar sind die einzelnen Segmente unterschiedlich breit, d. h. die Segmentierung ist nicht äquidistant. Dies ergibt sich aufgrund des Zentroidenverfahrens.

[0053] Im übrigen sei noch einmal angemerkt, daß das erfindungsgemäße Verfahren und der Mustererkenner nicht auf das Trainieren von Hidden-Markov-Modellen beschränkt, sondern prinzipiell bei jedem Mustererkennungsverfahren vorteilhaft anwendbar sind, bei denen eine Segmentierung von Trainingsmustern erforderlich ist.

[0054] Zusammenfassend ergeben sich durch die Erfindung im wesentlichen zwei Vorteile:

1. Es müssen nicht die Grenzen jedes einzelnen Trainingsmusters bestimmt werden. Statt dessen werden nur die Grenzen der Trainingsmuster nach dem Korrelieren und Verschieben bestimmt; diese Grenzen können dann auf alle zusammengehörenden Trainingsmuster angewendet werden.
2. Entsprechende Segmente verschiedener Trainingsmuster repräsentieren entsprechende Signalbereiche.

#### Patentansprüche

1. Verfahren zum Trainieren eines Modells für die Mustererkennung mittels mehrerer gleichartiger Trainingsmuster, insbesondere eines Hidden-Markov-Modells, bei dem
  - a) ein die Verschiebung zwischen mindestens zwei zusammengehörenden Trainingsmustern (11, 13) darstellender Verschiebungswert durch Korrelieren der zwei Trainingsmuster (11, 13) derart bestimmt wird, daß die Korrelation maximal wird, und
  - b) die Trainingsmuster (11, 13) mittels des Verschiebungswertes derart zueinander verschoben werden, daß einander entsprechende Bereiche der Trainingsmuster bezüglich eines gemeinsamen Bezugspunktes etwa die gleiche Lage aufweisen.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß das Korrelieren iterativ über mehrere Durchläufe für die Trainingsmuster (11, 13) durchgeführt wird.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß in einem Trainingsmuster (13) im Vergleich mit anderen zusammengehörenden Trainingsmustern (11) fehlende Bereiche durch vorgegebene Füllbereiche ergänzt werden.
4. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß anhand der in Schritt b) vorgenommenen Verschiebung Mustergrenzen (14, 16) bestimmt werden, die für alle zusammengehörenden Trainingsmuster (11, 13) gelten.
5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß ein zwischen den Mustergrenzen (14, 16) liegender Bereich der Trainingsmuster mittels eines Zentroidenverfahrens segmentiert wird.
6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, daß innerhalb vorgegebener Beschränkungen alle möglichen Segmentierungen berechnet werden und eine Segmentierung daraus nach einem Distanzkriterium zum Zentroiden jedes Segments ausgewählt wird.
7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß die Segmentierung ausgewählt wird, bei der die Summe der Quadrate der Distanzen von Merkmalsvektoren zu dem zugeordneten Zentroiden in jedem Segment über das gesamte Trainingsmuster minimal wird.
8. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Segmentierung entweder direkt zum Erstellen eines Modells zur Mustererkennung, insbesondere eines Hidden-Markov-Modells, oder als Initialsegmentierung für die Verschiebung von Segmentgrenzen (17, 18) einzelner Trainingsmuster, insbesondere mittels Viterbi-Segmentierung, dient.
9. Verfahren nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Mustererkennung eine Spracherkennung ist.
10. Mustererkenner zur Durchführung des Verfahrens nach einem der vorhergehenden Ansprüche, mit
- Vorverarbeitungsmitteln zum Vorverarbeiten eines ein Trainingsmuster repräsentierendes Signal,
  - Berechnungsmitteln zum Erstellen von Matrizen mit Merkmalsvektoren für das Trainingsmuster,
  - Korrelationsmitteln zum Korrelieren von Matrizen, die zusammengehörende Trainingsmuster repräsentieren,
  - Auswertemitteln zum Ermitteln eines Verschiebungswertes, und
  - Sortiermitteln zum Umsortieren von Matrizen abhängig von dem ermittelten Verschiebungswert.
11. Mustererkenner nach Anspruch 10, dadurch gekennzeichnet, daß Additionsmittel vorgesehen sind, welche die Matrizen nach einem Umsortieren addieren.
12. Mustererkenner nach Anspruch 10 oder 11, dadurch gekennzeichnet, daß er ein Spracherkenner ist.

---

Hierzu 3 Seite(n) Zeichnungen

---

